

ABSTRACT OF THE DISCLOSURE

Techniques for new event detection are provided. For a new story and a corpus of stories, story-pairs based on the new story and each corpus story are determined. Adjustments to the importance of terms are determined based on story characteristics associated with each story. Story characteristics are based on direct or indirect characteristics. Direct story characteristics include authorship, language associated with a story and the like. Indirect story characteristics may include derived characteristics such as an ROI category characteristic, a same ROI characteristic, a same event-same source characteristic, an average story similarity characteristic or any other known or later developed characteristic associated with a story. Adjustments to the inter-story similarity metrics are then determined based on story characteristics and/or a weighting function. New event scores and/or new event categorizations for stories are determined based on the inter-story similarity metrics and the adjustments based on the story characteristics. Optionally new events are selected based on new event scores and a threshold value.